Specification

Abstract

[0001] An increasing number of mobile telephones and computers are being equipped with a camera. Thus, instead of simple text strings, it is also possible to send images as queries to search engines or databases. Moreover, advances in image recognition allow a greater degree of automated recognition of objects, strings of letters, or symbols in digital images. This makes it possible to convert the graphical information into a symbolic format, for example, plain text, in order to then access information about the object shown.

Specification

[0002] A person sees an object and his or her memory immediately provides information related to the object. A system that emulates, or even expands upon, this ability would be extremely useful.

[0003] Modern image recognition processes allow ever-better recognition of objects, landscapes, faces, symbols, strings of letters, etc. in images. More and more cameras are connected to devices that are connected to remote data transmission networks. Such a configuration supports the following application. With the camera in a terminal (1), for example, in a mobile telephone, an image or a short sequence of images is recorded. This image (2) or these images are then sent to a server computer (7) by means of remote data transmission (3). In the server computer, an image recognition process (4) is run that converts the image information into symbolic information (5), for example, plain text. For example, the image recognition process may recognize that the Eiffel Tower can be seen in the image. The remaining process functions in a manner similar to a traditional search engine (6) on the Internet. The server computer sends the user back a list with "links" to database entries or web sites containing information about the object (8) shown.

1. Image Recognition

[0004] This section provides a rough overview of a possible method of object recognition. A more precise description of object recognition processes is described in the following publications: J. Buhmann, M. Lades, and C. v.d. Malsburg, "Size and Distortion Invariant Object recognition by Hierarchical Graph Matching," in Proceedings of the IJCNN International Joint Conference on Neural Networks, San Diego 1990, pages II-411-416 and "High-Level Vision: Object Recognition and Visual Cognition," Shimon Ullman, MIT Press; ISBN: 0262710072; July 31, 2000. Automatic character recognition processes are described in "Optical Character Recognition: An Illustrated Guide to the Frontier," Kluwer International Series in Engineering and Computer Science, 502, by Stephen V. Rice, George Nagy, Thomas A. Nartker, 1999.

1.1 Structure of an Object Representation

[0005] Most object recognition processes that are in use today use a number of example images (21) to train attribute detectors (22) adapted to the object.

1.2 Recognition

[0006] In recognition, the trained attribute detectors (32) are used to find the attributes they represent in an input image (31). This occurs by means of a search process. Each attribute detector outputs a confidence value that states how well it recognizes the attribute it represents from the image. If the accumulated confidence values (33) from all of the attribute detectors exceed a predetermined threshold value, it is assumed that the object was recognized.

2. Exemplary Embodiments

[0007] Naturally, automatic image recognition is still a long way from achieving the abilities of human vision. Therefore, we must first limit ourselves to situations that may be easily handled by existing image processing systems. In the following, I will describe a series of fields of application and discuss their specific difficulties.

City and Museum Guides

[0008] Visual recognition of buildings is easily attainable with today's methods. It is of course helpful if the user photographs the building in a frontal and vertical manner rather than from an oblique angle. Moreover, image recognition may be supported by using positioning information as well. Many telephones are equipped with GPS (Global Positioning System) such that it is possible to know the location of the telephone within a few meters at all times. This information can be used in image processing to limit consideration to only the buildings or building details that are nearby. Because the building ought to be recognizable at different times of the day, it is important to ensure when constructing the visual representation that appropriate image material must be incorporated. For most image recognition processes, this means that several pictures should be taken under various lighting conditions and these pictures should be used in constructing the models.

[0009] It would also be very simple to construct a universal art guide that would provide information about a painting, for example. Because pictures are two-dimensional, recognition is significantly simpler in this application.

Product Information

[0010] Another category of objects is products such as automobiles, books, or toys. If the user sees a model of automobile that he or she finds interesting, the user can simply take a picture of it and, for example, be pointed to a corresponding web site with more product information. Again, in the early phases of such a service, it will be useful if the user takes photographs from exact frontal or side views and sends them to the server computer. In later versions, when the pose invariance has been improved, the user will be less restricted. It is important for the image-based search service to be structured in such a way that, similar to the current World Wide Web, it is possible for every provider of information to offer an image-based search function for his or her web site. In this manner, it is possible to easily ensure that an image-based search function is available for many products because automobile manufacturers, for example, have a significant interest in their latest model being recognizable by imaging techniques.

Text Recognition

[0011] Another useful service lies in offering text recognition. For a person traveling to Tokyo or Paris who is not familiar with the local language, it would be very valuable to be able to point his or her camera at a sign and receive a translation and other information about the recognized text. If, for example, a person is standing in front of a sushi bar in Tokyo, it would be very valuable to be able to read the corresponding entry in a restaurant guide immediately and without additional effort. This is a particularly convenient solution to enable visitors who cannot read Japanese characters to access additional information.

Face Recognition

[0012] Face recognition is another special case. People who, for whatever reason, would like others to be able to find out more about them quickly may make images of their face available that could then be used in image recognition.

The Fully Constructed System

[0013] The list of application fields could be continued for a long time. Catalogs for antiquities and identification books for plants and animals could be made significantly more efficient using the system described above. Or a person could visualize part of a device for which he or she needs a replacement or more information. The person could simply take a picture and be quickly referred to its identification and manufacturer or to the corresponding section of a user manual. A system that provides additional information about advertising billboards is another application. In each of these cases, the user simply takes a picture of the object in question and sends it to the computer on which the image recognition system is running. The image recognition system sends corresponding symbolic information describing the object to the search engine, which finally selects the information that is sent back to the user.

[0014] In the completed stage of construction, a system results that could be equated with an extremely visual memory. Each object, each piece of text, each symbol, each face, and finally a